

Milan Hrga, mag. ing. comp.¹

RAČUNALNI VID

Stručni rad / Professional paper

UDK 004.42.004.1

Računalni vid je multidisciplinarno područje u kojem glavnu riječ vode znanstvenici računarstva i matematike. Znanstveni radovi iz tog područja su često potkrijepljeni samo matematičkim izvodima i škrtim grafičkim prikazima. U ovom radu, uz objašnjenje funkcioniranja računalnog vida, dan je detaljniji prikaz matematičkih izračuna, te su dopunjene grafički prikazi.

Ključne riječi: računalni vid, Haarove značajke, AdaBoost algoritam, integralna slika, kaskadni klasifikator.

1. Uvod

Krajem 20. stoljeća započeo je sve ubrzaniji razvoj elektronike, računala, a time i softvera koji podržavaju sve bolja računala. Danas je toliko brz razvoj novih tehnologija i softvera tako da se gotovo svaki dan pojave inovacije u informatičkom svijetu. „Prosječan“ programer više ne može pratiti sve trendove i smjerove razvoja programskih jezika, već se mora fokusirati na jednu domenu. Jedna od takvih domena, koja je nova, ali rapidno se razvija je „računalni vid“.

Računalni vid daje mogućnost računalu da „vidi“, to jest razumije što se nalazi na fotografiji ili videu. Računalni vid nam omogućuje razvoj potpuno novih aplikacija za svakodnevnu upotrebu. Primarna bi mu svrha mogla biti sigurnost i zaštita, kao što su: zaštita ljudi, zaštita podataka, vrijednih stvari, sprječavanje i reduciranje opasnosti od terorističkih napada i sl.

Uz pomoć računalnog vida uz relativno nisku ekonomsku potrošnju moguće je sniziti troškove, u odnosu na troškove, kad se koriste usluge ljudi, za obavljanje istih poslova. S daljnjim razvojem računalnog vida, mogao bi se smanjiti broj grešaka kod analiziranja određenih informacija, u odnosu na ljudski faktor. U početku implementacije u svakodnevnom životu, računalni vid bi mogao pomoći u rješavanju kriminalnih aktivnosti, detektirao bi ljude koji su počinili nekakvo nedjelo ili kriminalnu radnju. Ljudskom spoznajom postojanja određenog sustava nadzora smanjuje se učestalost kriminalnih aktivnosti. Svakodnevno se uočavaju nove potrebe za računalnim vidom u svakodnevnim aktivnostima u koje se može implementirati. Jedan od takvih je prepoznavanje kupaca i njihovih dotadašnjih navika i potreba u kupovini. Zbog pojedinaca koji su prepoznali mogućnosti koje donosi „računalni vid“, danas

¹ Veleučilište u Šibeniku

već postoje automobili koji sami voze, mobilni telefoni koji se mogu otključavati isključivo uz pomoć algoritma za prepoznavanje lica, te još mnoge inovativne implementacije u svakodnevnom životu. Još uvijek nismo svjesni koliko je ovaj segment jak i pogodan za razvoj i koristan za buduću sinergiju u suradnji računala/strojeva i ljudi kao u zajedničkom radu u proizvodnji, vožnji i sl.

Ovaj rad treba objasniti kako računoalo određuje postoji li na slici određena vrsta objekta, kao što je na primjer prometni znak, ljudsko lice, tumor na medicinskoj slici, i razno. Znanje iz računalnog vida je interdisciplinarno primjenjivo.

Računoalo može koristiti za računalni vid veće područje valnih duljina svjetlosti u odnosu na ljudsko oko. Glavni cilj rada je objasniti teorijsku pozadinu postupka otkrivanja objekata na slici, koristeći Viola Jones algoritam.

2. Računalni vid

Računalni vid znanstvena je i tehnološka disciplina koja se bavi teorijom i izradom samog sustava koji služe dobivanju informacija iz slika, bilo to iz jedne ili više fotografija, video uradaka ili određenih medicinskih uređaja. Uz to cilj računalnog vida je prepoznavanje objekata, praćenje objekata, detekcija unaprijed zadanih događaja, rekonstrukcija slike i sl.

Sam računalni vid srodan je mnogim drugim znanstvenim disciplinama s kojima je blisko vezan. Najbliži je umjetnoj inteligenciji, fotografiji i optici te je primjenjiv u mnogim disciplinama. Računalni vid je izrazito primjenjiv u proizvodnoj industriji za pozicioniranje i djelovanje robotskih ruku, uočavanje zastoja te uočavanje neispravnosti u proizvodnji, u medicini se implementira u velikom broju dijagnostičkih uređaja poput CT-a, u prometnim svrhama za autonomna vozila poput automobila, kamiona, koji se u današnje doba natječu tko će prvi pokrenuti proizvodnju komercijalnog vozila koje će u potpunosti autonomno moći prevoziti ljude ili teret bez ikakvog vozača. Uz sve to implementira se u svrhe sigurnosti, nadzora zračnih luka, skeniranja lica kako bi se prepoznale moguće prijetnje i opasnosti od terorizma te spriječile kriminalne radnje i u vojne svrhe za navođenje raketa, rekonstrukcija slika iz projekcija (CT), snimanja s velikih udaljenosti, geologija, poljoprivreda, nadzor i mjerenje proizvodnih procesa, očitavanje registracijskih pločica, čitanje bar koda, upravljanje robotom, samostalno kretanje vozila, očitavanje otisaka prstiju, analiza krvi i mnoge druge. Kako vrijeme prolazi, sve boljim razvojem računalnog vida implementacija i svrha će biti u sve širem spektru u svim svakodnevnim radnjama.

3. Algoritmi računalnog vida

3.1. SIFT

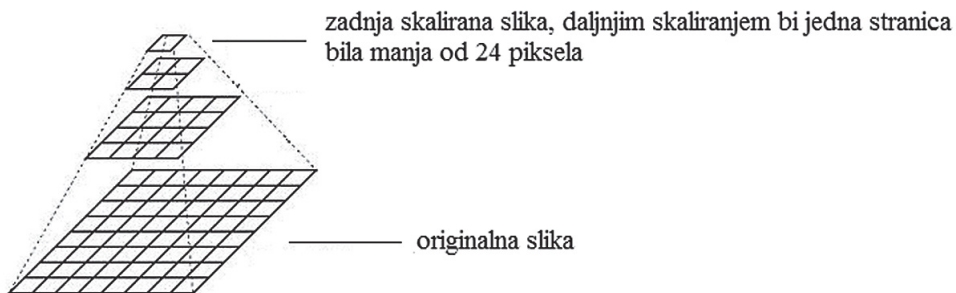
SIFT (*Scale-Invariant Feature Transform*) služi za pronalazak i opisivanje lokalnih značajki neke slike. Prednost ovog algoritma je invarijantnost s obzirom na rotaciju i uvećanje slike te relativna otpornost na promjene osvjetljenja i zaklonjenosti traženog objekta. SIFT ima velike mogućnosti kod trodimenzionalnog rekonstruiranja scene i uparivanja tj. prepoznavanja. Najvažnije svrhe su mu u robotici te spajanju više fotografija u jednu veću fotografiju. Zbog mogućnosti da relativno brzo i precizno usporedi trenutnu sliku s velikom bazom podataka

ima veliku mogućnost u nadgledanju, bilo u svrhu nadgledanja i traženja neželjenih objekata ili prebrojavanja (ljudi, automobila...) i sl.

3.2. Viola-Jones algoritam

Algoritam Viola-Jones primarno je napravljen za otkrivanje ljudskih lica na slici, ali može se koristiti za pronalaženje i bilo kakvih drugih objekata za koje ga se nauči kao npr. uočavanje pješaka, automobila, prometnih znakova, pješačkih prijelaza, oružja, dima(požara) i sl. Viola-Jones algoritam funkcionira tako da strojnim učenjem aplikaciji dajemo informacije da li svaka slika, od slika iz baze podataka, sadrži ili ne sadrži traženi objekt. Nakon dovršetka edukacije, prelazi se na sljedeću fazu, a to je kreiranje potprozora, najčešće dimenzija 24x24 piksela. Tim potprozorom prolazi se preko cijele slike, tako što pomičemo potprozor za određeni broj piksela i nastavljamo postupak dok ne dođemo do kraja ekrana. Dolaskom do kraja ekrana i završetkom obrade prelazi se na sljedeći korak kad se slika množi s koeficijentom k ($k < 1$) te ponavlja cijeli postupak dok se slika ne smanji do dimenzija manjih od 24 x 24 piksela, tj. dok u sliku više ne može stati potprozor veličine 24x24 piksela. Tada se postupak prekida. Svrha tih smanjivanja slike je da se lice osobe ili traženi predmet uklopi u jedan od potprozora dimenzija 24px*24px.

Slika 1. Primjer skaliranja veličine slike kada se množi koeficijentom k



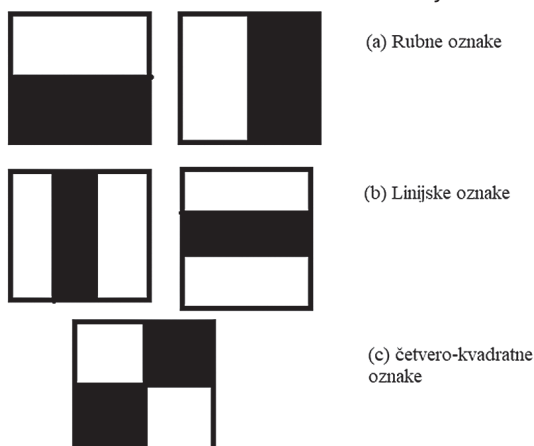
Izvor slike: itersnews.com/?p=99953

Algoritam Viola-Jones obuhvaća: Haarove značajke, integralnu sliku, Ada-Boost algoritam i kaskadni klasifikator.

3.3. Haarove značajke

Umjesto direktnog rada s pikselima koji bi nam otežavao posao, uveden je rad sa značajkama koje imaju pravokutni oblik. Na *slici 2* su osnovne Haarove značajke.

Slika 2. Osnovne Haarove značajke



Izvor: https://docs.opencv.org/trunk/d7/d8b/tutorial_py_face_detection.html

Skaliranjem osnovnih Haarovih značajki, na slici dimenzija 24px*24px dobijemo 45,396 [4] različitih značajki, kojih dio vidimo na slici 3.

Slika 3. Primjer Haarovih značajki na fotografiji



Izvor: <https://camo.githubusercontent.com/d128d742270321fb1aa48329584f6ff4d7adb02c/687474703a2f2f6661726d382e7374617469636666c696366b722e636f6d2f373230352f363838343335343632305f343935303362376333615f7a2e6a7067>

Testiranje svih značajki, na svakom potprozoru, bi dugo trajalo i samim tim bi bilo neučinkovito. Zbog toga se testiranja provode na skupu značajki koje su isfiltrirane u procesu treniranja, kao značajke koje su bitne u efikasnom pronalaženju određenog objekta. Izračun vrijednosti značajki olakšava nam postupak računanja integralne slike.

3.4. Integralna slika

Svaki element u integralnoj slici S_i dobije se tako da se izračunaju zbrojevi osvjetljenja svakog piksela originalne slike lijevo i iznad traženog slikovnog elementa, uključujući i traženi:

$$s(x,y) = s(x,y-1) + i(x,y)$$

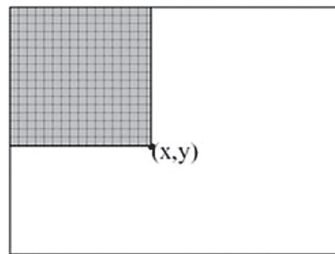
$$ii(x,y) = ii(x-1,y) + s(x,y)$$

$i(x,y)$ je osvjetljenje piksela na poziciji x,y

$s(x,y)$ je suma osvjetljenja piksela u retku, do pozicije x,y pri čemu je uključena i pozicija x,y

$ii(x,y)$ je suma svih piksela iznad i lijevo od pozicije x,y , uključujući i poziciju x,y (Slika 4)

Slika 4. Element integralne slike na koordinatama x i y računa se kao suma slikovnih elemenata u označenom pravokutniku



Izvor: www.cognotics.com/opencv/servo_2007_series/part_2/sidebar.html

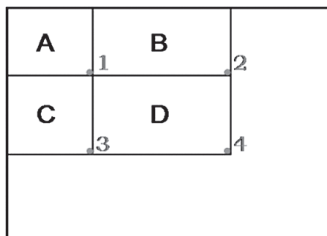
Ovom metodom se vrlo brzo izračuna suma slikovnih elemenata unutar bilo kojeg pravokutnika u originalnoj slici kao što je ilustrirano na slici 5.

Za izračun Haarove značajke s dva pravokutnika, dovoljno je pristupiti integralnoj slici šest puta, za značajku s tri pravokutnika, osam pristupa, za značajku s 4 pravokutnika samo 9 puta.

Suma slikovnih elemenata u pravokutniku D (slika 5) može se izračunati kao:

$D = S_{i4} - S_{i3} - S_{i2} + S_{i1}$, gdje su $S_{i1}, S_{i2}, S_{i3}, S_{i4}$ vrijednosti elemenata u integralnoj slici S_i na označenim lokacijama. To se može lako pokazati uvrštavanjem $S_{i1} = A$, $S_{i2} = A+B$, $S_{i3} = A+C$, $S_{i4} = A+B+C+D$. Pojednostavljeno rečeno postoji analogija s računanjem površine. Površina D pravokutnika je jednaka površini iznad i lijevo od točke 4, minus površina gore i lijevo od točke 2 (površina $A+B$), minus površina gore i lijevo od točke 3 (površina $C+A$), te plus površina gore i lijevo od točke 1 (površinu A smo prethodno 2 puta oduzeli).

Slika 5. Suma slikovnih elemenata u pravokutniku D



Izvor: www.cognotics.com/opencv/servo_2007_series/part_2/sidebar.html

3.5. AdaBoost algoritam

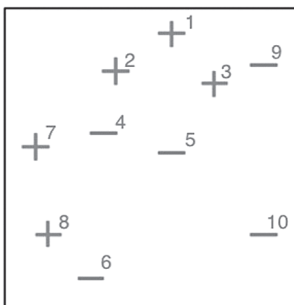
AdaBoost algoritam² obavlja funkciju poboljšanja performansi jednostavnog algoritma strojnog učenja. Ideja algoritma je stvaranje jednog jakog klasifikatora od više slabih klasifikatora. Slabi klasifikator je funkcija koja točno klasificira više od 50% uzoraka. Jaki klasifikator se dobiva linearnom kombinacijom slabih klasifikatora (slika 7). Matematička definicija jakog klasifikatora opisana je sljedećom relacijom:³

$$H(x) = \sum_{n=1}^N a_n h_n(x)$$

gdje $H(x)$ predstavlja jaki klasifikator, a $h(x)$ slabi klasifikator.

U sljedećem primjeru je prikazan način, pomoću slika i tablice, na koji se gradi jaki klasifikator, za otkrivanje lica na slici, uz pomoć inicijalnog ljudskog određivanja, to jest učenja računala da li na nekoj slici postoji ili ne postoji lice. U nastavku je grafički, numerički i tablični prikaz dobijanja parametara koji služe za dobijanje jakog klasifikatora. U skupu od 10 fotografija dimenzija 24px * 24px na 5 fotografija se nalazi lice, a na 5 se ne nalazi lice. Oznaka „+“ (na slici 6.1. označeno brojevima 1-10) označava one fotografije iz skupa u kojima se nalazi lice, a oznaka „-“ označava da se na promatranoj fotografiji ne nalazi lice.

Slika 6.1. Distribucija uzoraka



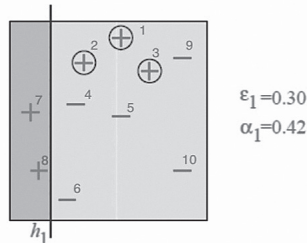
Izvor: learning.cis.upenn.edu/cis520_fall2009/index.php?n=lectures.boosting

² Meir, R., and Rätsch, G. (2003). An introduction to boosting and leveraging. In *Advanced lectures on machine learning*, Springer, Berlin, Heidelberg.

³ Kim, T. K., i Cipolla, R. (2013). Multiple classifier boosting and tree-structured classifiers. In *Machine Learning for Computer Vision*, Springer, Berlin, Heidelberg.

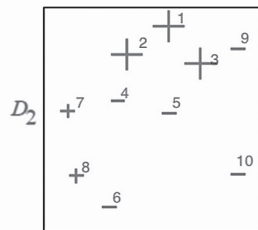
Slabim klasifikatorom⁴ „ h_1 “ suklasificirane slike pri čemu su, iz skupa fotografija, pogrešno klasificirane fotografije označene oznakama 1, 2 i 3 na slici 6.1. U slijedećem koraku potreban je onaj slabi klasifikator koji bolje klasificirati one uzorke koje je prethodni slabi klasifikator loše klasificirao (slika 6.2). Povećavamo težine onim fotografijama (slike 6.2, 6.3) koje je neispravno klasificirao klasifikator „ h_1 “ i normaliziramo numeričke vrijednosti. (Tablica 1, slika 9).

Slika 6.2. Klasifikacija fotografija pomoću klasifikatora „ h_1 “



Izvor: learning.cis.upenn.edu/cis520_fall2009/index.php?n=lectures.boosting

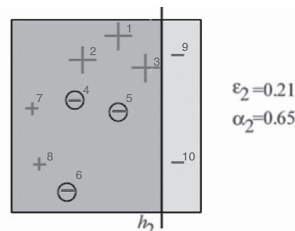
Slika 6.3. Povećanje težina neispravno i smanjenje težina ispravno klasificiranih fotografija u skupu nakon klasificiranja klasifikatora „ h_1 “



Izvor: learning.cis.upenn.edu/cis520_fall2009/index.php?n=lectures.boosting

Slabi klasifikatorom „ h_2 “ (slika 6.4.) pogrešno je klasificirao fotografije 4, 5 i 6 kao da se na njima nalazi lice, a to su fotografije na kojima se ne nalazi lice, dok je ostale fotografije ispravno klasificirao.

Slika 6.4. Klasifikator „ h_2 “

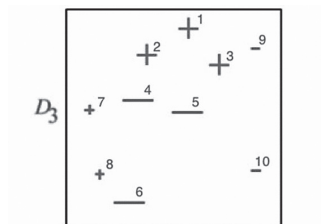


Izvor: learning.cis.upenn.edu/cis520_fall2009/index.php?n=lectures.boosting

⁴ Padilla, R., Costa Filho, C. F. F., i Costa, M. G. F. (2012). Evaluation of haar cascade classifiers designed for face detection. World Academy of Science, Engineering and Technology.

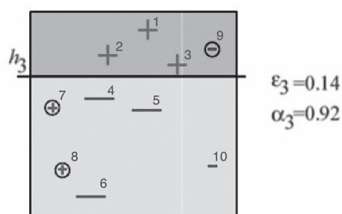
Povećavaju se težine uzorcima koje je klasifikator „h2“ neispravno klasificirao.

Slika 6.5. Povećanje težina neispravno i smanjenje težina ispravno klasificiranih fotografija u skupu nakon klasificiranja klasifikatora „h2“



Izvor: learning.cis.upenn.edu/cis520_fall2009/index.php?n=lectures.boosting

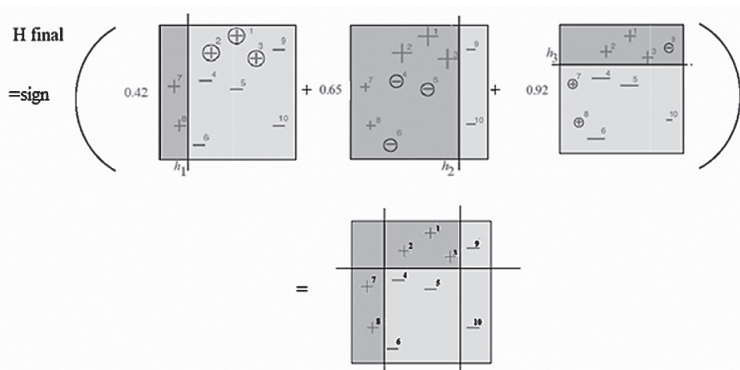
Slika 6.6. Klasifikacija fotografija pomoću klasifikatora „h3“



Izvor: learning.cis.upenn.edu/cis520_fall2009/index.php?n=lectures.boosting

Linearnom kombinacijom slabih klasifikatora „h1“, „h2“ i „h3“ dobije se jaki klasifikator $H(x)$ ⁵ (slika 7).

Slika 7. Jaki klasifikator



Izvor: learning.cis.upenn.edu/cis520_fall2009/index.php?n=lectures.boosting

⁵ Meir, R., and Rätsch, G. (2003). An introduction to boosting and leveraging. In Advanced lectures on machine learning, Springer, Berlin, Heidelberg.

Tablica 1. Povećanje težina neispravno klasificiranim fotografijama i smanjenje težina ispravno klasificiranim (eng.boosting)

D_i	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	Z_i
$D_1(i)$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0,9165
$D_2(i)$	0.167	0.167	0.167	0.0714	0.0714	0.0714	0.0714	0.0714	0.0714	0.0714	0,8207
$D_3(i)$	0.1061	0.1061	0.1061	0.16675	0.1667	0.1667	0.0455	0.0455	0.0455	0.0455	0,6614
$D_4(i)$	0.0637	0.0637	0.0637	0.1001	0.1001	0.1001	0.1730	0.1730	0.1730	0,0273	-

Izvor: autorov izračun.

Slika 8. Isječak koda u Matlabu za pojašnjenje normalizacije distribucije

```

X= [-1,-1,-1,1,1,1,1,1,1,1];
% X(i)=yi*hc(xi)
Z=0;
for i=1:10
    D(i)=D(i)*exp(-alpha*X(i));
    Z=Z+D(i);
end

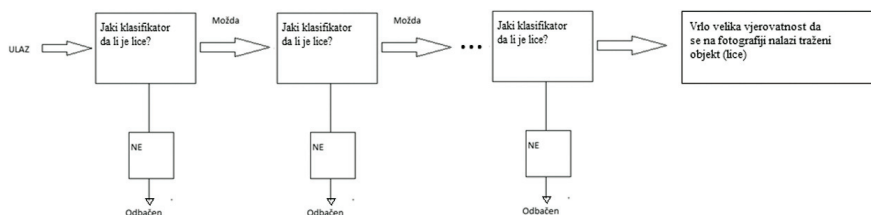
sum=0;
for i=1:10
    D(i)=D(i)/Z;
    display(D(i));
    sum=sum+Dn(i);%provjera, suma D0(i)mora biti jednaka 1
end
    
```

Izvor: autor.

3.6. Kaskadni klasifikator

Ako je bilo koji jaki klasifikator, iz kaskade⁶ odbacio mogućnost da se na fotografiji nalazi lice, postupak se prekida, to jest na fotografiji (24px * 24 px) se ne nalazi lice (slika 9).

Slika 9. Kaskadni klasifikator



Izvor:autor.

⁶ Viola, P. i Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. I-II). IEEE.

4. Zaključak

Računalni vid, iako još nije dosegao zadovoljavajuće razine sigurnosti, već u određenim segmentima mogu služiti, barem kao pomoć ljudskom radu. Uz pomoć računalnog vida kod npr. traženja određenih osoba preko sigurnosnih kamera, trenutno je moguće barem suziti izbor na manje od 10% detektiranih osoba na kamerama, što znatno olakšava ljudski rad. Također u autonomnoj vožnji samo je pitanje vremena kada ćemo moći u potpunosti prepustiti vožnju autonomnim vozilima bez da time ugrozimo ljudske živote i okoliš. Tehnologija je sve bliže (rezolucije kamera, brzine računala) proizvodnji potpuno sigurnih vozila ili rakete na navođenje koja neće ugroziti živote civila nego samo eliminirati prijetnju (opasnost). Daljnji napredak je moguć primjenom kombiniranih tehnologija poput elektroničkog nosa i IC kamera, te analizom ljudskih pokreta poput hoda. Iako napredak ove znanstvene grane ide sporije od predviđenog, donosi znatne pogodnosti za lakši tehnološki napredak počevši od same proizvodnje pa do svakodnevne upotrebe u vidu bilo kakve pomoći ljudima i do veće sigurnosti.

LITERATURA

1. Kim, T. K., i Cipolla, R. (2013). Multiple classifier boosting and tree-structured classifiers. In *Machine Learning for Computer Vision*, Springer, Berlin, Heidelberg.
2. Meir, R., and Rätsch, G. (2003). An introduction to boosting and leveraging. In *Advanced lectures on machine learning*, Springer, Berlin, Heidelberg.
3. Padilla, R., Costa Filho, C. F. F., i Costa, M. G. F. (2012). Evaluation of haar cascade classifiers designed for face detection. *World Academy of Science, Engineering and Technology*.
4. Viola, P., i Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-I). IEEE.
5. Wikipedia //en.wikipedia.org/wiki/Scale-invariant_feature_transform

Summary

COMPUTER VISION

Computer vision is a multidisciplinary scientific field which is discussed and understood by mathematicians and computing professionals. Scientific papers about the mentioned scientific field are usually described by mathematical formulas and graphs, and not explained into detail, which would enable a better understanding to scientists from other scientific fields. This paper tries to give a more detailed explanation of computer vision, but it contains mathematical calculations and graphs as well.

Keywords: computer vision, Haar features, AdaBoost algorithm, integral image, cascade classifier.